BRIEF COMMUNICATION

A Study of the Impact of Adding HPV Types to Cervical Cancer Screening and Triage Tests

Mark Schiffman, Michelle J. Khan, Diane Solomon, Rolando Herrero, Sholom Wacholder, Allan Hildesheim, Ana Cecilia Rodriguez, Maria C. Bratti, Cosette M. Wheeler, Robert D. Burk

For the PEG Group and the ALTS Group

Use of human papillomavirus (HPV) testing in cervical cancer prevention is increasing rapidly. A DNA test for 13 HPV types that can cause cervical cancer is approved in the United States for co-screening with cytology of women ≥30 years old and for triage of women of all ages with equivocal cytology. However, most infections with HPV are benign. We evaluated trade-offs between specificity and sensitivity for approximately 40 HPV types in predicting cervical intraepithelial neoplasia 3 and cancer in two prospective studies: a population-based screening study that followed 6196 women aged 30–94 years from Costa Rica for 7 years and a triage study that followed 3363 women aged 18–90 years with equivocal cytology in four U.S. centers for 2 years. For both screening and triage, testing for more than about 10 HPV types decreased specificity more than it increased sensitivity. The minimal increases in sensitivity and in negative predictive value achieved by adding HPV types to DNA tests must be weighed against the projected burden to thousands of women falsely labeled as being at high risk of cervical cancer. [J Natl Cancer Inst 2005;97:147–50]

Human papillomavirus (HPV) DNA testing has been shown in numerous epidemiologic studies worldwide to increase the early detection of cervical cancer and its precursors when used alone or with cytology screening *(1–3).* Because early detection and treatment can greatly reduce cervical cancer incidence and mortality, the U.S. Food and Drug Administration (FDA) recently approved HPV DNA testing as an adjunct to Papanicolaou screening in women aged 30 years and over and for triage of women of all ages with equivocal cytology *(3–6).* Several HPV test systems exist; so far, the FDA has approved one test that targets 13 HPV genotypes that are known to cause cervical cancer (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) *(7).*

A recent report suggested that five HPV types that are not included in the FDA-approved probe set (HPV26, 53, 66, 73, and 82) cause small percentages of cervical cancers *(8),* raising the question of whether next-generation HPV tests should include these HPV types. In countries where screening test sensitivity is viewed as much more important than specificity, the tendency is to probe for all HPV types that are linked to either cancer or cervical intraepithelial neoplasia grade 3 (CIN3), the immediate precursor to cancer. However, the specificity of HPV DNA testing is a major concern now that millions of women worldwide are receiving such tests, with a sizable proportion testing positive and therefore being sent for colposcopy and possible treatment. Pooled HPV probe sets yield overall positive versus negative, rather than type-specific, results; therefore, categorizing as oncogenic those HPV types that are only occasionally associated with cancer could adversely influence clinical judgment and patient management, resulting in substantial costs, patient anxiety, and iatrogenic morbidity. Moreover, colposcopy itself has mediocre accuracy *(9)* and excisional treatment of CIN has been linked to an increased risk of premature births in subsequent pregnancies due to the rupture of membranes *(10).*

To evaluate the impact of modifying the HPV types in a probe set, we analyzed data from two large studies conducted by the U.S. National Cancer Institute. The Proyecto Epidemiológico Guanacaste (PEG) was a randomly sampled, prospective screening study carried out in a high-risk province in Costa Rica *(11).* The ASCUS-LSIL Triage Study (ALTS) was a randomized clinical trial on the management of equivocal and low-grade squamous intraepithelial lesions (LSIL) carried out in four U.S. centers *(12,13).*

The designs of PEG and ALTS have been published previously *(11,13).* For this analysis, we included only women who would be recommended for HPV testing following current U.S. guidelines *(4–6)*, i.e., women ≥30 years old in a screening setting (as studied in PEG) and women of all ages with equivocal cytologic interpretations called atypical squamous cells of undetermined significance (ASCUS; as studied in ALTS). Specifically, HPV test results and outcome data were available for 6196 (99.6%) of 6223 women aged 30–94 years screened in PEG (mean follow-up, 6.3 ± 1.3 years) and 3363 (96.4%) of 3488 women aged 18–90 years (mean follow-up, 1.8 ± 0.6 years) studied in ALTS. We assayed more than 40 HPV types by using a TaqGold polymerase with dot blot hybridization of MY09/MY11 polymerase chain reaction (PCR) amplification products in PEG and 38 HPV types with linear array hybridization of PGMY PCR products in ALTS *(14,15).*

To account for the insensitivity of colposcopy as a reference standard *(13,16,17),* we incorporated results from clinical follow-up in defining CIN3 and cancer. We reasoned that enrollment screening tests should identify both

Affiliations of authors: Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, MD (MS, MJK, SW, AH); Howard Hughes Medical Institute, Chevy Chase, MD (MJK); Division of Cancer Prevention, NCI, NIH, DHHS, Bethesda, MD (DS); Proyecto Epidemiológico Guanacaste (PEG), Guanacaste, Costa Rica (RH, ACR, MCB); Departments of Molecular Genetics and Microbiology and Obstetrics and Gynecology, University of New Mexico Health Sciences Center, School of Medicine, Albuquerque, NM (CMW); Albert Einstein College of Medicine, Bronx, NY (RDB).

Correspondence to: Mark Schiffman, MD, MPH, Division of Cancer Epidemiology and Genetics, NCI, 6120 Executive Blvd. Room 7066, EPS MSC 7234, Bethesda, MD 20892-7234 (e-mail: schiffmm@mail.nih.gov).

See "Notes" following "References."

DOI: 10.1093/jnci/dji014

Journal of the National Cancer Institute, Vol. 97, No. 2,

prevalent and incipient CIN3—in total, all cases diagnosed within 2 years *(17)*. For PEG screening, we included CIN3 within 2 years of enrollment (n = 61) and all cervical cancers diagnosed throughout the study (n = 37) *(17)*. For ALTS triage of ASCUS, we included CIN3 (n = 291) and cervical cancers (n = 2) diagnosed within the 2-year trial *(13)*.

For PEG and ALTS separately, we estimated how test performance would vary depending on which HPV types were targeted. We iteratively added types in order of maximum impact on sensitivity to determine which types would maximize test sensitivity, and we evaluated changes in sensitivity and specificity for each addition. We plotted the receiver operating characteristic (ROC) curves for each population and allowed the data to guide the results without consideration of which types are in the currently approved assay. HPV16 was the first type to be included for both study populations because, of all types, it was present in the largest number of CIN3 and cancers in both populations and therefore had maximum impact on sensitivity. After HPV16, the next HPV types found in the largest number of CIN3 and cancer samples were HPV58 in PEG and HPV31 in ALTS. We continued iterations for each population for up to 20 HPV types; because the last additions did not improve sensitivity, these types were added in order of the smallest decrease in specificity.

For screening of women ≥30 years in PEG (Fig. 1), the percentage decrease in specificity was larger than the percentage increase in sensitivity after the probe pool included 12 HPV types (HPV16, 58, 18, 31, 56, 51, 11, 68, 52, 35, 45, and 66) and no additional sensitivity was gained after inclusion of the thirteenth type (HPV71), which detected one case. With these 13 probes, the sensitivity was 89.8% (95% confidence interval [CI] = 82.0% to 95.0%) and the specificity was 88.8% (95% CI = 88.0% to 89.6%). By comparison, using the PCR data for the 13-type combination included in the current FDA-approved probe set showed a sensitivity of 86.7% (95% CI = 78.4% to 92.7%) and a specificity of 90.4% (95% CI = 89.6% to 91.1%). HPV33 and HPV59, which are included in the FDA-approved kit, did not come up in our iterative model (perhaps due to random effects reflecting the small numbers of these types in this population). HPV11, which is associated with anogenital and laryngeal warts, was found in two multiply infected cases of CIN3 (one with an oncogenic type, HPV33).

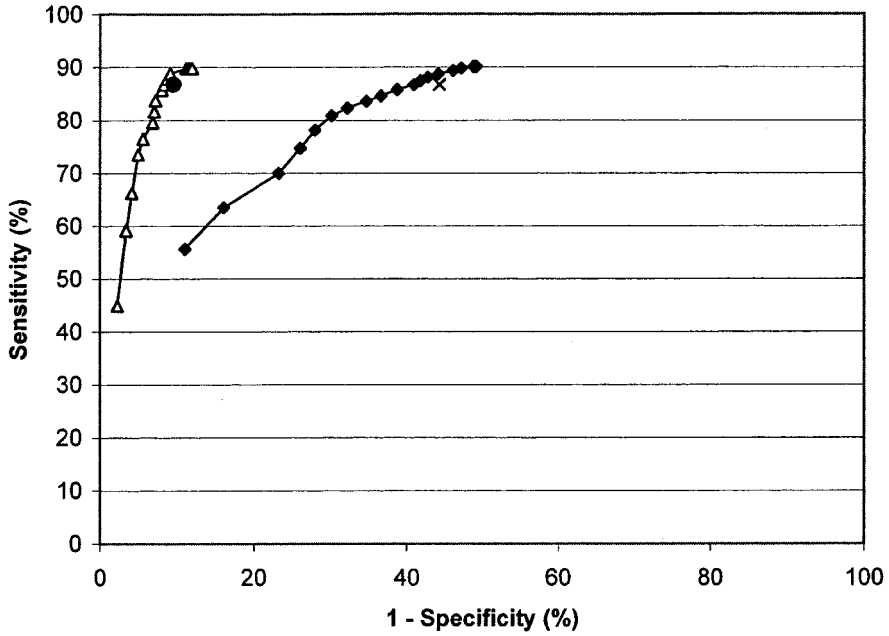

Fig. 1. Receiver operating characteristic (ROC) curve for human papillomavirus (HPV) types in a screening or triage test. Sensitivities of detection of cervical intraepithelial neoplasia grade 3 (CIN3) and cervical cancer were plotted versus (100% – specificity) for increasingly sensitive combinations of HPV types. The results for screening in the Proyecto Epidemiológico Guanacaste (PEG) are plotted with **open triangles**, and the results for triage of equivocal cytology in the ASCUS-LSIL Triage Study (ALTS) are plotted with **solid diamonds.** As described in the text, up to 20 HPV types were added in order of maximum impact on sensitivity and, when the impact on sensitivity was the same, in order of minimum decrement in specificity. For PEG, HPV types were added in the following order: HPV16, 58, 18, 31, 56, 51, 11, 68, 52, 35, 45, 66, 71, 67, 74, 59, 40, 32, 55, and 89. For ALTS, HPV types were added in the following order: HPV16, 31, 52, 58, 33, 35, 45, 18, 42, 66, 51, 73, 82, 54, 39, 84, 53, 57, 11, and 26. The last several HPV types added in each study did not increase sensitivity, and their corresponding points are indistinguishable on the plots. For each study, we show as a separate point (**solid circle** for PEG, **cross** for ALTS) the estimated performance of a probe set containing the 13 HPV types in the currently FDA-approved HPV DNA test.

Specificity of HPV testing was much lower for triage of ASCUS cases in ALTS (Fig. 1) than it was for screening in PEG because overall HPV prevalence in ASCUS (including infections destined to clear) is higher than that in a screening population. Specificity decreased more than sensitivity increased after eight HPV types were added (HPV16, 31, 52, 58, 33, 35, 45, and 18); at this iteration, sensitivity was 80.9% (95% CI = 75.9% to 85.2%) and specificity was 69.8% (95% CI = 68.2% to 71.5%). Marginal (and perhaps some random) increases in sensitivity continued until the seventeenth type was added (HPV16, 31, 52, 58, 33, 35, 45, 18, 42, 66, 51, 73, 82, 54, 39, 84, and 53), at which point sensitivity was 90.1% (95% CI = 86.1% to 93.3%). However, at this point, specificity was only 51.3% (95% CI = 49.5% to 53.1%). PCR data simulating the FDA-approved 13-type pool yielded a sensitivity of 86.7% (95% CI = 82.3% to 90.4%) and a specificity of 55.7% (95% CI = 53.9% to 57.5%). HPV56, -59, and -68 (which are in the FDA-approved kit) did not appear in our iterations because they occurred only in multiple infections with HPV types that contributed more to sensitivity.

Figure 2 demonstrates the projected population impact of adding HPV types to screening and triage tests by plotting true-positive results (numbers of HPV-positive women referred for colposcopy who would in fact have CIN3 or cancer) versus false-positive results (numbers of HPV-positive women referred for colposcopy who would not have CIN3 or cancer, i.e., a measure of the burden of testing). To make these projections, we used the sensitivity and specificity for each of the HPV combinations shown in Fig. 1 and assumed a hypothetical screen-

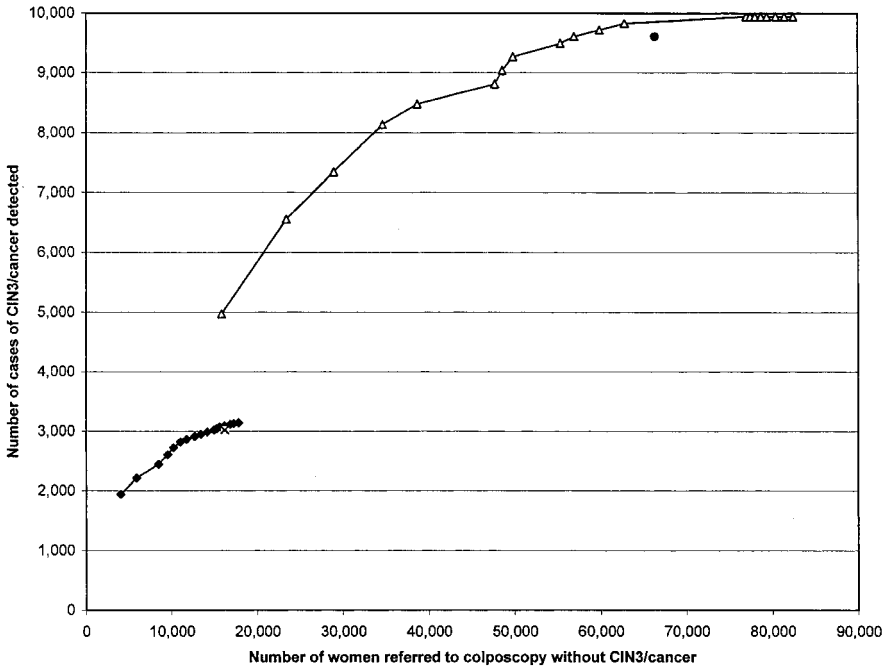

Fig. 2. Projected population impact of adding types to a human papillomavirus (HPV) screening or triage test. For the same HPV combinations described in Fig. 1, and in the same order, the number of cases of cervical intraepithelial neoplasia grade 3 (CIN3) and cervical cancer that would be detected by increasingly sensitive combinations of HPV types was plotted against the number of women referred to colposcopy with a positive HPV test that would not have CIN3 or cancer (true positives versus false positives). We chose a hypothetical screening population of 1 000 000 women with screening and triage results corresponding to those in the Proyecto Epidemiologico Guanacaste (PEG) and the ASCUS-LSIL Triage Study (ALTS), respectively. To project the impact of each HPV test combination on screening, we assumed that 70% of women (700 000) would be ≥30 years of age and screened with cytology and adjunctive HPV testing. To project the impact on triage, we assumed that 4% (40 000) of women of all ages would have ASCUS triaged by HPV testing. The results for screening in PEG are plotted with **open triangles**, and the results for triage of equivocal cytology in ALTS are plotted with **solid diamonds.** For each study, we show as a separate point (**solid circle** for PEG, **cross** for ALTS) the estimated performance of a probe set containing the 13 HPV types targeted by the currently FDA-approved HPV DNA test. It should be noted that triage is shown at bottom left and screening at upper right, showing that screening is the setting where choice of HPV types is particularly critical.

ing population of 1 000 000 women. To project the impact of each HPV test combination on screening, we assumed that 70% of women in the population (700 000) would be ≥30 years of age and screened with cytology and adjunctive HPV testing *(18).* To project the impact on triage of ASCUS, we assumed that 4% (40 000) of women of all ages would have ASCUS and would be triaged by HPV testing *(19).* We found that, for both screening and triage, adding HPV types beyond the 13-type kit would lead to very few additional detected cases but would result in thousands of additional women without CIN3 or cancer being unnecessarily referred to colposcopy. The absolute numbers in Fig. 2 demonstrate that the health burden (i.e., unnecessary referrals for further follow-up) resulting from decreased specificity is particularly great in a screening setting.

In summary, our results show that approximately 90% of prevalent and incipient CIN3 and cancer cases in both screening and triage settings could be detected by a single enrollment HPV test, leading to very high negative predictive values. However, the optimal types were not exactly the same for the Costa Rican and U.S. populations. Some HPV types are indisputably important for an HPV test kit that will be used in screening and/or triage, but adding all possibly carcinogenic types will inevitably produce a nonspecific test.

Our study populations were large and representative, permitting a detailed examination of trade-offs between sensitivity and specificity. However, we had fewer CIN3 and cancer cases than the largest international series of cervical cancer *(20),* which remain the best epidemiologic guides regarding which HPV types can cause cancer. Nonetheless, even large international series can result in debatable conclusions regarding the correct types to include in screening tests. For example, in the largest case control study published to date, HPV53 was designated as a probable carcinogen because a single case (of 1918 cancers) contained HPV53 alone *(8).* However, HPV53 was one of the most common HPV types among women without CIN3 or cancer in our two populations (data not shown). We can estimate with reasonable precision that adding HPV53 to the current FDA-approved 13-probe pool would statistically significantly decrease screening specificity by 1.7% (95% CI = 1.4% to 2.1%, $P<.001$ by McNemar's test). In absolute numbers, testing 700 000 women aged ≥30 years old for HPV53 would result in approximately 10 500 women unnecessarily targeted for intensified clinical management with virtually no additional detection of CIN3 and cancer.

The present analysis demonstrates testing trade-offs but is not intended to replace formal cost–utility analyses. However, lowered specificity, resultant unnecessarily intensified clinical management, and iatrogenic morbidity should be seriously considered by those developing and judging the next generation of HPV DNA assays, particularly for those tests that are intended to be applied to screening of the general population.

REFERENCES

(1) Arbyn M, Buntinx F, Van Ranst M, Paraskevaidis E, Martin-Hirsch P, Dillner J. Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst 2004; 96:280–93.

(2) Cuzick J, Szarewski A, Cubie H, Hulman G, Kitchener H, Luesley D, et al. Management of women who test positive for high-risk types of human papillomavirus: the HART study. Lancet 2003; 362:1871–6.

(3) Wright TC, Jr., Schiffman M. Adding a test for human papillomavirus DNA to cervical-cancer screening. N Engl J Med 2003; 348: 489–90.

(4) Saslow D, Runowicz CD, Solomon D, Moscicki AB, Smith RA, Eyre HJ, Cohen C, for the American Cancer Society. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. CA Cancer J Clin 2002; 52:342–62.

(5) Wright TC, Jr., Cox JT, Massad LS, Twiggs LB, Wilkinson EJ, for the ASCCP-Sponsored Consensus Conference. 2001 Consensus Guidelines for the management of

women with cervical cytological abnormalities. JAMA 2002; 287:2120–9.
(6) ACOG practice bulletin: clinical management guidelines for obstetrician-gynecologists. Number 45, August 2003. Cervical cytology screening (replaces committee opinion 152, March 1995). Obstet Gynecol 2003; 102:417–27.
(7) Wright TC, Jr., Schiffman M, Solomon D, Cox JT, Garcia F, Goldie S, et al. Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening. Obstet Gynecol 2004; 103:304–9.
(8) Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ, for the International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 2003; 348:518–27.
(9) Guido R, Schiffman M, Solomon D, Burke L, for the ASCUS LSIL Triage Study (ALTS) Group. Postcolposcopy management strategies for women referred with low-grade squamous intraepithelial lesions or human papillomavirus DNA-positive atypical squamous cells of undetermined significance: a two-year prospective study. Am J Obstet Gynecol 2003; 188:1401–5.
(10) Sadler L, Saftlas A, Wang W, Exeter M, Whittaker J, McCowan L. Treatment for cervical intraepithelial neoplasia and risk of preterm delivery. JAMA 2004; 291:2100–6.
(11) Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, Morales J, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J Natl Cancer Inst 2000; 92:464–74.
(12) Solomon D, Schiffman M, Tarone R, for the ALTS Group. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. J Natl Cancer Inst 2001; 93:293–9.
(13) ASCUS LSIL Triage Study (ALTS) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 2003; 188:1383–92.
(14) Castle PE, Schiffman M, Gravitt PE, Kendall H, Fishman S, Dong H, et al. Comparisons of HPV DNA detection by MY09/11 PCR methods. J Med Virol 2002; 68:417–23.
(15) Peyton CL, Gravitt PE, Hunt WC, Hundley RS, Zhao M, Apple RJ, Wheeler CM. Determinants of genital human papillomavirus detection in a US population. J Infect Dis 2001; 183:1554–64.
(16) Cox JT, Schiffman M, Solomon D, for the ASCUS LSIL Triage Study (ALTS) Group. Prospective follow-up suggests similar risk of subsequent cervical intraepithelial neoplasia grade 2 or 3 among women with cervical intraepithelial neoplasia grade 1 or negative colposcopy and directed biopsy. Am J Obstet Gynecol 2003; 188:1406–12.
(17) Ferreccio C, Bratti MC, Sherman ME, Herrero R, Wacholder S, Hildesheim A, et al. A comparison of single and combined visual, cytologic, and virologic tests as screening strategies in a region at high risk of cervical cancer. Cancer Epidemiol Biomarkers Prev 2003; 12:815–23.
(18) Insinga RP, Glass AG, Rush BB. Pap screening in a U.S. health plan. Cancer Epidemiol Biomarkers Prev 2004; 13:355–60.
(19) Davey DD, Woodhouse S, Styer P, Stastny J, Mody D. Atypical epithelial cells and specimen adequacy: current laboratory practices of participants in the college of American pathologists interlaboratory comparison program in cervicovaginal cytology. Arch Pathol Lab Med 2000; 124:203–11.
(20) Clifford GM, Smith JS, Plummer M, Munoz N, Franceschi S. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. Br J Cancer 2003; 88:63–73.

NOTES

The following individuals participated in the PEG Group: M. Schiffman, project officer, A. Hildesheim, S. Wacholder, statistician, M. Sherman, pathologist, José Jeronimo, gynecologic oncologist, at the National Cancer Institute, NIH, Bethesda MD; M.C. Bratti, principal investigator, R. Herrero, co-principal investigator, A. C. Rodriguez, assistant principal investigator, J. Morales, gynecologic oncologist and colposcopist, M. Alfaro, cytopathologist, and D. Guillen, histopathologist, at the Costa Rican Foundation for Health Sciences, San Jose, Costa Rica; M. Hutchinson, cytologist, at Women and Infants' Hospital, Providence, RI; and J. Schussler, senior programmer and analyst, at Information Management Services, Inc., Silver Spring, MD.

The following individuals participated in the ALTS Group: D. Solomon, project officer, M. Schiffman, co-project officer, and P. Castle, National Cancer Institute, NIH, Bethesda, MD; at clinical centers: E.E. Partridge, principal investigator, University of Alabama at Birmingham, AL, J.L. Walker, principal investigator, G.A. Johnson, co-principal investigator, and A. Yadack, study manager, University of Oklahoma, Oklahoma City, OK: R.S. Guido, principal investigator, K. McIntyre-Seltman, co-principal investigator, R.P. Edwards, investigator, and J. Gruss, study manager, Magee-Womens Hospital of the University of Pittsburgh Medical Center Health System, Pittsburgh, PA, as part of the Colposcopy Quality Control Group: D. Ferris, principal investigator, Medical College of Georgia, Augusta, GA; J.T. Cox, co-investigator, University of California at Santa Barbara, Santa Barbara, CA: L. Burke, co-investigator, Beth Israel Deaconess Medical Center Hospital, Boston, MA; HPV Quality Control Group: C.M. Wheeler, principal investigator, and C. Peyton-Goodall, lab manager, University of New Mexico Health Sciences Center, Albuquerque, NM, and M.M. Manos, co-investigator, Kaiser Permanente, Oakland, CA; Pathology Quality Control Group: R.J. Kurman, principal investigator, and D.L. Rosenthal, co-investigator, Johns Hopkins Hospital, Baltimore, MD; M.E. Sherman, co-investigator, National Cancer Institute, NIH, Rockville, MD; M.H. Stoler, co-investigator, University of Virginia Health Science Center, Charlottesville, VA; at Westat, the coordinating unit, Rockville, MD: J. Rosenthal, project director; M. Dunn, data management team leader; and J. Quarantillo, senior systems analyst; Quality of Life Group: D. M. Harper, chair of ALTS Quality of Life Group, Dartmouth Medical School, Hanover, NH; at Digene Corporation, Gaithersburg, MD: A.T. Lorincz, senior scientific officer; and at Information Management Services, Inc., Silver Spring, MD, B. Kramer, senior programmer and analyst.

PEG was supported by the National Cancer Institute, NIH DHHS contracts NO1-CP-21081, NO1-CP-33061, NO1-CP-40542, and NO1-CP-506535. Dr. Burk was supported by an NIH grant. We thank Dr. Thomas Wright for histologic review of enrollment cases.

ALTS was supported by the National Cancer Institute, NIH DHHS contracts CN-55153, CN-55154, CN-55155, CN-55156, CN-55157, CN-55158, CN-55159, and CN-55105.

Some of the equipment and supplies used in these studies were donated or provided at reduced cost by Digene Corporation, Gaithersburg, MD; Cytyc Corporation, Boxborough, MA; National Testing Laboratories, Fenton, MO; Denvu, Tucson, AZ; TriPath Imaging, Inc., Burlington, NC; and Roche Molecular Systems, Alameda, CA.

Manuscript received July 7, 2004; revised November 3, 2004; accepted November 8, 2004.